SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Donson, Jon

Dawson, William 0. Grantham, George L. Turpen, Thomas H.

Turpen, Ann Myers Garger, Stephen J. Grill, Laurence K.

- (ii) TITLE OF INVENTION: RECOMBINANT PLANT VIRAL NUCLEIC ACIDS
- (iii) NUMBER OF SEQUENCES: 11
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Limbach & Limbach
 - (B) STREET: 2001 Ferry Building
 - (C) CITY: San Francisco
 - (D) STATE: CAL
 - (F) ZIP: 94111
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: Patent in Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 923,692
 - (B) FILING DATE: 31-JUL-1992
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 600,244
 - (B) FILING DATE: 22-OCT-1990
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 641,617
 - (B) FILING DATE: 16-JAN-1991
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 310,881
 - (B) FILING DATE: 17-FEB-1989
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 160,766
 - (B) FILING DATE: 26-FEB-1988
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 160,771
 - (B) FILING DATE: 26-FEB-1988
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 347,637
 - (B) FILING DATE: 05-MAY-1989
- (vii) PRIOR APPLICATION DATA:

CERTIFICATE OF MAILING

i hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Commissioner of Patents and Trademarks,

Washington, DC 20231 on_

4/28/93

Dated: 4/28/93

| SEQ.L97 | T - April 26, 1993 | Page 1-2 |
|-----------|--|----------|
| - · | (A) APPLICATION NUMBER: US (B) FILING DATE: 08-JUN-198 | |
| (vii) | PRIOR APPLICATION DATA: (A) APPLICATION NUMBER: US (B) FILING DATE: 15-JUL-19 | |
| (viii) | ATTORNEY/AGENT INFORMATION: (A) NAME: Halluin, Albert 1 (B) REGISTRATION NUMBER: 28 (C) REFERENCE/DOCKET NUMBER | 8,957 |
| (ix) | TELECOMMUNICATION INFORMATION (A) TELEPHONE: 415-433-4150 (B) TELEFAX: 415-433-8716 | |
| (2) II | NFORMATION FOR SEQ ID NO: 1: | |
| (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 4 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear | |
| (ii) | MOLECULE TYPE: peptide | |
| (iii) | HYPOTHETICAL: NO | |
| (iv) | ANTI-SENSE: NO | |
| (xi) | SEQUENCE DESCRIPTION: SEQ IN | D NO: 1: |
| P1 . 1 | ro Xaa Gly Pro | |
| (2) I | NFORMATION FOR SEQ ID NO: 2: | |
| (i) | SEQUENCE CHARACTERISTICS: (A) LENGTH: 13 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| (ii) | MOLECULE TYPE: DNA (genomic |) |
| (iii) | HYPOTHETICAL: NO | |
| (iv) | ANTI-SENSE: NO | |
| (xi) | SEQUENCE DESCRIPTION: SEQ I | D NO: 2: |
| GGGTAC | CTGG GCC | |
| | | |

- (2) INFORMATION FOR SEQ ID NO: 3:
 - (i) SEQUENCE CHARACTERISTICS:

| SEQ.LST ~ | . Apri | 1 26, 1993 | | Page 1-3 | | |
|------------------|-------------------------|---|---------------|--------------------|---------------|------------|
| - | (B) TYPE: (C) STRAN | H: 886 base nucleic ac DEDNESS: si OGY: linear | id | i d | | |
| (ii) | MOLECULE T | YPE: DNA (g | enomic) | | | |
| (iii) | HYPOTHETIC | AL: NO | | | | |
| (iv) | ANTI-SENSE | : NO | | | | |
| (vi) | ORIGINAL S (A) ORGAN | OURCE: ISM: Chines | e cucumbe | er · | | |
| (vii) | IMMEDIATE (B) CLONE | SOURCE: : alpha-tri | chosanthi | .n | | |
| (ix) | | KEY: CDS (B | | ON: 8877 | | |
| (xi) | SEQUENCE D | ESCRIPTION: | SEQ ID N | TO: 3: | | |
| CTCGAGG AT | IG ATC AGA | TTC TTA GTC | CTC TCT | TTG CTA ATT | CTC ACC | CTC 49 |
| Me | et Ile Arg 1 | Phe Leu Val 5 | Leu Ser | Leu Leu Ile 10 | Leu Thr | Leu |
| TTC CTA AC | CA ACT CCT | GCT GTG GAG | GGC GAT | GTT AGC TTC | CGT TTA | TCA 97 |
| Phe Leu Th 15 | hr Thr Pro | Ala Val Glu 20 | Gly Asp | Val Ser Phe 25 | Arg Leu | Ser 30 |
| GGT GCA A | CA AGC AGT | TCC TAT GGA | GTT TTC | ATT TCA AAT | CTG AGA | AAA 145 |
| Gly Ala Th | hr Ser Ser 35 | Ser Tyr Gly | Val Phe 40 | Ile Ser Asn | Leu Arg 45 | Lys |
| GCT CTT C | CA AAT GAA | AGG AAA CTG | TAC GAT | ATC CCT CTG | TTA CGT | TCC 193 |
| Ala Leu Pi | ro Asn Glu 50 | Arg Lys Leu | Tyr Asp 55 | Ile Pro Leu | Leu Arg 60 | Ser |
| TCT CTT C | CA GGT TCT | CAA CGC TAC | GCA TTG | ATC CAT CTC | ACA AAT | TAC 241 |
| | ro Gly Ser 65 | Gln Arg Tyr 70 | Ala Leu | Ile His Leu 75 | Thr Asn | Tyr |
| GCC GAT GA | AA ACC ATT | TCA GTG GCC | ATA GAC | GTA ACG AAC | GTC TAT | ATT 289 |
| Ala Asp G 80 | lu Thr Ile | Ser Val Ala 85 | Ile Asp | Val Thr Asn 90 | Val Tyr | Ile |
| ATG GGA TA | AT CGC GCT | GGC GAT ACA | TCC TAT | TTT TTC AAC | GAG GCT | TCT 337 |
| Met Gly Ty 95 | yr Arg Ala | Gly Asp Thr 100 | Ser Tyr | Phe Phe Asn 105 | | Ser 110 |

| SEQ.LST | u. | April | 26. 1 | 993 | | | Page | e 1-4 | | | | | |
|-------------|----------------|---------------|--------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-----|
| GÇA ACA | GAA GCT | - | • | | ጥጥር | ΔΔΔ | | , | | CCA | ΔΔΔ | ርጥጥ | 385 |
| Ala Thr | | | | | | | | | | | | • | 303 |
| " | GIU AIA | 115 | · | vai | FIIE | 120 | Asp | на | Mec | Arg | 125 | vai | |
| ACG CTT | CCA TAT | TCT GG | CAAT | TAC | GAA | AGG | CTT | CAA | ACT | GCT | GCG | GGC | 433 |
| Thr Leu | Pro Tyr 130 | Ser Gl | y Asn | Tyr | Glu 135 | Arg | Leu | Gln | Thr | Ala 140 | Ala | Gly | |
| AAA ATA | AGG GAA | AAT AT | r ccg | CTT | GGA | CTC | CCA | GCT | TTG | GAC | AGT | GCC | 481 |
| Lys Ile | Arg Glu 145 | Asn Il | e Pro | Leu 150 | Gly | Leu | Pro | Ala | Leu 155 | Asp | Ser | Ala | |
| ATT ACC | ACT TTG | TTT TA | C TAC | AAC | GCC | AAT | TCT | GCT | GCG | TCG | GCA | CTT | 529 |
| Ile Thr | Thr Leu | Phe Ty | Tyr 165 | Asn | Ala | Asn | Ser | Ala 170 | Ala | Ser | Ala | Leu | |
| ATG GTA | CTC ATT | CAG TC | ACG | TCT | GAG | GCT | GCG | AGG | TAT | AAA | TTT | ATT | 577 |
| Met Val : | Leu Ile | Gln Se | | Ser | Glu | Ala 185 | Ala | Arg | Tyr | Lys | Phe | Ile 190 | |
| GAG CAA | CAA ATT | GGG AA | G CGC | GTT | GAC | AAA | ACC | TTC | СТА | CCA | AGT | TTA | 625 |
| Glu Gln | Gln Ile | Gly Ly 195 | a Arg | Val | Asp | Lys 200 | Thr | Phe | Leu | Pro | Ser 205 | Leu | |
| GCA ATT | ATA AGT | TTG GA | AAT | AGT | TGG | TCT | GCT | CTC | TCC | AAG | CAA | ATT | 673 |
| Ala Ile | Ile Ser 210 | Leu Gl | ı Asn | Ser | Trp 215 | Ser | Ala | Leu | Ser | Lys 220 | Gln | Ile | |
| CAG ATA | GCG AGT | ACT AA | TAA 1 | GGA | CAG | TTT | GAA | ACT | CCT | GTT | GTG | CTT | 721 |
| Gln Ile | Ala Ser 225 | Thr As | ı Asn | Gly 230 | Gln | Phe | Glu | Thr | Pro 235 | Val | Val | Leu | |
| ATA AAT | GCT CAA | AAC CA | A CGA | GTC | ATG | ΑŢΑ | ACC | AAT | GTT | GAT | GCT | GGA | 769 |
| Ile Asn 240 | Ala Gln | Asn Gl | 1 Arg 245 | Val | Met | Ile | Thr | Asn 250 | Val | Asp | Ala | Gly | |
| GTT GTA | ACC TCC | AAC AT | C GCG | TTG | CTG | CTG | AAT | CGA | AAC | AAT | ATG | GCA | 817 |
| Val Val | Thr Ser | Asn Il 26 | | Leu | Leu | Leu | Asn 265 | Arg | Asn | Asn | Met | Ala 270 | |
| GCC ATG | GAT GAC | GAT GT | г сст | ATG | ACA | CAG | AGC | TTT | GGA | TGT | GGA | AGT | 865 |
| Ala Met | Asp Asp | Asp Va 275 | l Pro | Met | Thr | Gln 280 | Ser | Phe | Gly | Cys | Gly 285 | Ser | |
| TAT GCT | ATT TAG | FAACTCG | AG | | | | | | | | | | 886 |
| Tyr Ala | Ile | | | | | | | | | | | | |

. 290

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 289 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu Phe Leu 1 5 10 15

Thr Thr Pro Ala Val Glu Gly Asp Val Ser Phe Arg Leu Ser Gly Ala
20 25 30

Thr Ser Ser Tyr Gly Val Phe Ile Ser Asn Leu Arg Lys Ala Leu 35 40 45

Pro Asn Glu Arg Lys Leu Tyr Asp Ile Pro Leu Leu Arg Ser Ser Leu 50 55 60

Pro Gly Ser Gln Arg Tyr Ala Leu Ile His Leu Thr Asn Tyr Ala Asp 65 70 75 80

Glu Thr Ile Ser Val Ala Ile Asp Val Thr Asn Val Tyr Ile Met Gly
85 90 95

Tyr Arg Ala Gly Asp Thr Ser Tyr Phe Phe Asn Glu Ala Ser Ala Thr 100 105 110

Glu Ala Ala Lys Tyr Val Phe Lys Asp Ala Met Arg Lys Val Thr Leu 115 120 125

Pro Tyr Ser Gly Asn Tyr Glu Arg Leu Gln Thr Ala Ala Gly Lys Ile 130 135 140

Arg Glu Asn Ile Pro Leu Gly Leu Pro Ala Leu Asp Ser Ala Ile Thr 145 150 155 160

Thr Leu Phe Tyr Tyr Asn Ala Asn Ser Ala Ala Ser Ala Leu Met Val 165 170 175

Leu Ile Gln Ser Thr Ser Glu Ala Ala Arg Tyr Lys Phe Ile Glu Gln
180 185 190

Gln Ile Gly Lys Arg Val Asp Lys Thr Phe Leu Pro Ser Leu Ala Ile 195 200 205

Ile Ser Leu Glu Asn Ser Trp Ser Ala Leu Ser Lys Gln Ile Gln Ile 210 215 220

Ala Ser Thr Asn Asn Gly Gln Phe Glu Thr Pro Val Val Leu Ile Asn

| | <u> </u> |
|--|-----------------------------------|
| | |
| SEQ.LST April 26, 1993 | Page 1-6 |
| 225 230 | 235 240 |
| Ala Gln Asn Gln Arg Val Met Ile Thr Asn 245 250 | |
| Thr Ser Asn Ile Ala Leu Leu Leu Asn Arg 260 265 | Asn Asn Met Ala Ala Met 270 |
| Asp Asp Val Pro Met Thr Gln Ser Phe | e Gly Cys Gly Ser Tyr Ala 285 |
| Ile | |
| (2) INFORMATION FOR SEQ ID NO: 5: | |
| (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1450 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear | |
| (ii) MOLECULE TYPE: DNA (genomic) | • |
| (iii) HYPOTHETICAL: NO | |
| (iv) ANTI-SENSE: NO | |
| <pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Oryza sativa</pre> | |
| <pre>(vii)</pre> | |
| (ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATI (B) LOCATION: 121316 | CON: 121316 |
| (xi) SEQUENCE DESCRIPTION: SEQ ID | NO: 5: |
| CCTCGAGGTG C ATG CAG GTG CTG AAC ACC AT | TG GTG AAC A CAC TTC TTG 48 |
| Met Gln Val Leu Asn Thr Me 1 5 | et Val Asn Lys His Phe Leu 10 |
| TCC CTT TCG GTC CTC ATC GTC CTT GGC | C CTC TCC TCC AAC TTG ACA 96 |
| Ser Leu Ser Val Leu Ile Val Leu Leu Gly | Leu Ser Ser Asn Leu Thr 25 |
| GCC GGG CAA GTC CTG TTT CAG GGA TTC AAC | C TGG GAG TCG TGG AAG GAG 144 |
| Ala Gly Gln Val Leu Phe Gln Gly Phe Asr 30 35 | Trp Glu Ser Trp Lys Glu 40 .45 |

AAT GGC GGG TGG TAC AAC TTC CTG ATG GGC AAG GTG GAC GAC ATC GCC

Asn Gly Gly Trp Tyr Asn Phe Leu Met Gly Lys Val Asp Asp Ile Ala

| SEQ. | LST | | | Apri | L1 26 | , 15 | 93 | | | Page | € T-7 | / | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-----|
| | | | | 50 | | | | | 55 | 1 | ű | | | 60 | | , |
| GCA | GCC | GGC | ATC | ACC | CAC | GTC | TGG | CTC | CCT | CCG | CCG | TCT | CAC | TCT | GTC | 240 |
| Aĺa | Ala | Gly | Ile 65 | Thr | His | Val | Trp | Leu 70 | Pro | Pro | ,Pro | Ser | His 75 | Ser | Val | |
| GGC | GAG | CAA | GGC | TAC | ATG | CCT | GGG | CGG | CTG | TAC | GAT | CTG | GAC | GCG | TCT | 288 |
| Gly | Glu | Gln 80 | Gly | Tyr | Met | Pro | Gly 85 | Arg | Leu | Tyr | Asp | Leu 90 | Asp | Ala | Ser | |
| AAG | TAC | GGC | AAC | GAG | GCG | CAG | CTC | AAG | TCG | CTG | ATC | GAG | GCG | TTC | CAT | 336 |
| Lys | Tyr 95 | Gly | Asn | Glu | Ala | Gln 100 | Leu | Lys | Ser | Leu | Ile 105 | Glu | Ala | Phe | His | |
| GGC | AAG | GGC | GTC | CAG | GTG | ATC | GCC | GAC | ATC | GTC | ATC | AAC | CAC | CGC | ACG | 384 |
| Gly 110 | Lys | Gly | Val | Gln | Val 115 | Ile | Ala | Asp | Ile | Val 120 | Ile | Asn | His | Arg | Thr 125 | |
| GCG | GAG | CAC | AAG | GAC | GGC | CGC | GGC | ATC | TAC | TGC | CTC | TTC | GAG | GGC | GGG | 432 |
| Ala | Glu | His | Lys | Asp 130 | Gly | Arg | Gly | Ile | Tyr 135 | Cys | Leu | Phe | Glu | Gly 140 | Gly | |
| ACG | CCC | GAC | TCC | CGC | CTC | GAC | TGG | GGC | CCG | CAC | ATG | ATC | TGC | CGC | GAC | 480 |
| Thr | Pro | Asp | Ser 145 | Arg | Leu | Asp | Trp | Gly 150 | Pro | His | Met | Ile | Cys 155 | Arg | Asp | |
| GAC | CCC | TAC | GGC | CAT | GGC | ACC | GGC | AAC | CCG | GAC | ACC | GGC | GCC | GAC | TTC | 528 |
| Asp | Pro | Tyr 160 | | Asp | Gly | Thr | Gly 165 | Asn | Pro | Asp | Thr | Gly 170 | Ala | Asp | Phe | |
| GCC | GCC | GCG | CCG | GAC | ATC | GAC | CAC | CTC | AAC | AAG | CGC | GTC | CAG | CGG | GAG | 576 |
| Ala | Ala 175 | Ala | Pro | Asp | Ile | Asp 180 | His | Leu | Asn | Lys | Arg 185 | Val | Gln | Arg | Glu | |
| CTC | ATT | GGC | TGG | CTC | GAC | TGG | CTC | AAG | ATG | GAC | ATC | GGC | TTC | GAC | GCG | 624 |
| Leu 190 | Ile | Gly | Trp | Leu | Asp 195 | Trp | Leu | Lys | Met | Asp 200 | Ile | Gly | Phe | Asp | Ala 205 | · |
| TGG | CGC | CTC | GAC | TTC | GCC | AAG | GGC | TAC | TCC | GCC | GAC | ATG | GCA | AAC | ATC | 672 |
| Trp | Arg | Leu | Asp | Phe 210 | Ala | Lys | Gly | Tyr | Ser 215 | Ala | Asp | Met | Ala | Lys 220 | Ile | |
| TAC | ATC | GAC | GCC | ACC | GAG | CCG | AGC | TTC | GCC | GTG | CCC | GAG | ATA | TCG | ACG | 720 |
| Tyr | Ile | Asp | Ala 225 | Thr | Glu | Pro | Ser | Phe 230 | Ala | Val | Ala | Glu | Ile 235 | Trp | Thr | |
| TCC | ATG | GCG | AAC | GGC | GGG | GAC | GGC | AAG | CCG | AAC | TAC | GAC | CAG | AAC | GCG | 768 |

| Ser | Met | Ala 240 | Asn | Gly | Gly | Asp | Gly 245 | Lys | Pro | Asn | Tyr | Asp 250 | Gln | Asn | Ala | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------|
| CAC | CGG | CAG | GAG | CTG | GTC | AAC | TGG | GTC | GAT | CGT | GTC | GGC | GGC | GCC | AAC | 816 |
| His | Arg 255 | Gln | Glu | Leu | Val | Asn 260 | Trp | Val | Asp | Arg | Val 265 | Gly | Gly | Ala | Asn | |
| ACC | AAC | GGC | ACG | GCG | TTC | GAC | TTC | ACC | ACC | AAG | GGC | ATC | CTC | AAC | GTC | 864 |
| Ser 270 | Asn | Gly | Thr | Ala | Phe 275 | Asp | Phe | Thr | Thr | Lys 280 | Gly | Ile | Leu | Asn | Val 285 | |
| GCC | GTG | GAG | GGC | GAG | CTG | TGG | CGC | CTC | CGC | GGC | GAG | GAC | GGC | AAG | GCG | 912 |
| Ala | Val | Glu | Gly | Glu 290 | Leu | Trp | Arg | Leu | Arg 295 | Gly | Glu | Asp | Gly | Lys 300 | Ala | |
| CCC | GGC | ATG | ATC | GGG | TGC | TGG | CCG | GCC | AAG | GCG | ACG | ACC | TTC | GTC | GAC | 960 |
| Pro | Gly | Met | Ile 305 | Gly | Trp | Trp | Pro | Ala 310 | Lys | Ala | Thr | Thr | Phe 315 | Val | Asp | |
| AAC | CAC | GAC | ACC | GGC | TCG | ACG | CAG | CAC | CTG | TGG | CCG | TTC | CCC | TCC | GAC | 1008 |
| Asn | His | Asp 320 | Thr | Gly | Ser | Thr | Gln 325 | His | Leu | Trp | Pro | Phe 330 | Pro | Ser | Asp | |
| AAG | GTC | ATG | CAG | GGC | TAC | GCA | TAC | ATC | CTC | ACC | CAC | CCC | GGC | AAC | CCA | 1056 |
| Lys | Val 335 | Met | Gln | Gly | Tyr | Ala 340 | Tyr | Ile | Leu | Thr | His 345 | Pro | Gly | Asn | Pro | |
| TGC | ATC | TTG | TAC | GAC | CAT | TTC | TTC | GAT | TGG | GGT | CTC | AAG | GAG | GAG | ATC | 1104 |
| Cys 350 | Ile | Phe | Tyr | Asp | His 355 | Phe | Phe | Asp | Trp | Gly 360 | Leu | Lys | Glu | Glu | Ile 365 | |
| GAG | CGC | CTG | GTG | TCA | ATC | AGA | AAC | CGG | CAG | GGG | ATC | CAC | CCG | GCG | AGC | 1152 |
| Glu | Arg | Leu | Val | Ser 370 | Ile | Arg | Asn | Arg | Gln 375 | Gly | Ile | His | Pro | Ala 380 | Ser | |
| GAG | CTG | CGC | ATC | ATG | GAA | GCT | GAC | AGC | GAT | CTC | TAC | CTC | GCG | GAG | ATC | 1200 |
| Glu | Leu | Arg | Ile 385 | Met | Glu | Ala | Asp | Ser 390 | Asp | Leu | Tyr | Leu | Ala 395 | Glu | Ile | |
| GAT | GGC | AAG | GTG | ATC | ACA | AAG | ATT | GGA | CCA | AGA | TAC | GAC | GTC | GAA | CAC | 1248 |
| Asp | Gly | Lys 400 | Val | Ile | Thr | Lys | Ile 405 | Gly | Pro | Arg | Tyr | Asp 410 | Val | Glu | His | |
| CTC | ATC | CCC | GAA | GGC | TTC | CAG | GTC | GTC | GCG | CAC | GGT | GAT | GGC | TAC | GCA | 1296 |
| Leu | Ile 415 | Pro | Glu | Gly | Phe | Gln 420 | Val | Val | Ala | His | Gly 425 | Asp | Gly | Tyr | Ala | |

| April | 26, | 1993 |
|-------|-----|------|
| | , | |

SEQ.LST

Page 1-9

1351

1411

| ATC TGG GAG AAA ATC TGAGCGCACG ATGACGAGAC TCTCAGTTTA GCAGATTTAA | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Ile Trp Glu Lys LIe 430 435 | | | | | | | | | | | | | | |
| CCTGCGATTT TTACCCTGAC CGGTATACGT ATATACGTGC CGGCAACGAG CTGTATCCGA | | | | | | | | | | | | | | |
| TCCGAATTAC GGATGCAATT GTCCACGAAG TCCTCGAGG | | | | | | | | | | | | | | |
| (2) INFORMATION FOR SEQ ID NO: 6: | | | | | | | | | | | | | | |
| (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 434 amino acids (B) TYPE: amino acid (D) Topology: linear | | | | | | | | | | | | | | |
| (ii) MOLECULE TYPE: protein | | | | | | | | | | | | | | |
| (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6: | | | | | | | | | | | | | | |
| Met Gln Val Leu Asn Thr Met Val Asn Lys His Phe Leu Ser Leu Ser 1 5 10 15 | | | | | | | | | | | | | | |
| Val Leu Ile Val Leu Leu Gly Leu Ser Ser Asn Leu Thr Ala Gly Gln 20 25 30 | | | | | | | | | | | | | | |
| Val Leu Phe Gln Gly Phe Asn Trp Glu Ser Trp Lys Glu Asn Gly Gly 35 40 45 | | | | | | | | | | | | | | |
| Trp Tyr Asn Phe Leu Met Gly Lys Val Asp Asp Ile Ala Ala Gly 50 55 60 | | | | | | | | | | | | | | |
| Ile Thr His Val Trp Leu Pro Pro Pro Ser His Ser Val Gly Glu Gln 65 70 75 80 | | | | | | | | | | | | | | |
| Gly Tyr Met Pro Gly Arg Leu Tyr Asp Leu Asp Ala Ser Lys Tyr Gly 85 90 95 | | | | | | | | | | | | | | |
| Asn Glu Ala Gln Leu Lys Ser Leu Ile Glu Ala Phe His Gly Lys Gly 100 105 110 | | | | | | | | | | | | | | |
| Val Gln Val Ile Ala Asp Ile Val Ile Asn His Arg Thr Ala Glu His 115 120 125 | | | | | | | | | | | | | | |
| Lys Asp Gly Arg Gly Ile Tyr Cys Leu Phe Glu Gly Gly Thr Pro Asp 130 135 140 | | | | | | | | | | | | | | |
| Ser Arg Leu Asp Trp Gly Pro His Met Ile Cys Arg Asp Asp Pro Tyr 145 150 150 160 | | | | | | | | | | | | | | |
| Gly Asp Gly Thr Gly Asn Pro Asp Thr Gly Ala Asp Phe Ala Ala Ala 165 170 175 | | | | | | | | | | | | | | |
| Pro Asp Ile Asp His Leu Asn Lys Arg Val Gln Arg Glu Leu Ile Gly | | | | | | | | | | | | | | |

180

185

190

Trp Leu Asp Trp Leu Lys Met Asp Ile Gly Phe Asp Ala Trp Arg Leu 195 200 205

Asp Phe Ala Lys Gly Tyr Ser Ala Asp Met Ala Lys Ile Tyr Ile Asp 210 215 220

Ala Thr Glu Pro Ser Phe Ala Val Ala Glu Ile Trp Thr Ser Met Ala 225 230 235 240

Asn Gly Gly Asp Gly Lys Pro Asn Tyr Asp Gln Asn Ala His Arg Gln 245 250 255

Glu Leu Val Asn Trp Val Asp Arg Val Gly Gly Ala Asn Ser Asn Gly
260 265 270

Thr Ala Phe Asp Phe Thr Thr Lys Gly Ile Leu Asn Val Ala Val Glu 275 280 285

Gly Glu Leu Trp Arg Leu Arg Gly Glu Asp Gly Lys Ala Pro Gly Met 290 295 300

Ile Gly Trp Trp Pro Ala Lys Ala Thr Thr Phe Val Asp Asn His Asp 305 310 315 320

Thr Gly Ser Thr Gln His Leu Trp Pro Phe Pro Ser Asp Lys Val Met 325 330 335

Gln Gly Tyr Ala Tyr Ile Leu Thr His Pro Gly Asn Pro Cys Ile Phe 340 345 350

Tyr Asp His Phe Phe Asp Trp Gly Leu Lys Glu Glu Ile Glu Arg Leu 355 360 365

Val Ser Ile Arg Asn Arg Gln Gly Ile His Pro Ala Ser Glu Leu Arg 370 375 380

Ile Met Glu Ala Asp Ser Asp Leu Tyr Leu Ala Glu Ile Asp Gly Lys 385 390 395 400

Val Ile Thr Lys Ile Gly Pro Arg Tyr Asp Val Glu His Leu Ile Pro
405 410 415

Glu Gly Phe Gln Val Val Ala His Gly Asp Gly Tyr Ala Ile Trp Glu 420 425 430

Lys Ile

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 709 base pairs
 - (B) TYPE: nucleic acid
 - (G) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

| | April 26, 1993 Page 1-11 |
|-----------------|--|
| | MOLECULE TYPE: cDNA to mRNA |
| (iii) | HYPOTHETICAL: NO |
| (iv) | ANTI-SENSE: NO |
| (vi) | ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens |
| (vii) | IMMEDIATE SOURCE: (B) CLONE: alpha-hemoglobin |
| (ix) | FEATURE: (A) NAME/KEY: transit_peptide (B) LOCATION: 26241 (B) LOCATION: 26241 |
| (ix) | FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 245670 |
| (xi) | SEQUENCE DESCRIPTION: SEQ ID NO: 7: |
| CTCGAGGGC | A TCTGATCTTT CAAGAATGGC ACAAATTAAC AACATGGCAC AAGGGATACA 60 |
| AACCCTTAA | T CCCAATTCCA ATTTCCATAA ACCCCAAGTT CCTAAATCTT CAAGTTTTCT 120 |
| TGTTTTTGG | A TGTAAAAAAC TGAAAATTC AGCAAATTCT ATGTTGGTTT TGAAAAAAGA 180 |
| TTCAATTTT | T ATGCAAAAGT TTTGTTCCTT TAGGATTTCA GCAGGTGGTA GAGTTTCTTG 240 |
| CATG GTG | CTG TCT CCT GCC GAC AAG ACC AAC GTC AAG GCC GCC TGG GGC 289 |
| Vál 1 | Leu Ser Pro Ala Asp Lys Thr Asn Val Lys Ala Ala Trp Cly 5 10 15 |
| AAG GTT G | GC GCG CAC GCT GGC GAG TAT GGT GCG GAG GCC CTG GAG AGG 337 |
| Lys Val G | ly Ala His Ala Gly Glu Tyr Gly Ala Glu Ala Leu Glu Arg 20 25 30 |
| ATG TTC C | TG TCC TTC CCC ACC ACC AAG ACC TAC TTC CCG CAC TTC GAC 385 |
| Met Phe L | eu Ser Phe Pro Thr Thr Lys Thr Tyr Phe Pro His Phe Asp 35 40 45 |
| CTG AGC C | AC GGC TCT GCC CAG GTT AAG GGC CAC GGC AAG AAG GTG GCC 433 |
| Leu Ser H | is Gly Ser Ala Gln Val Lys Gly His Gly Lys Lys Val Ala |
| a.a.a.a. | 50 55 60 |
| | TG ACC AAC GCC GTG GCG CAC GTG GAC GAC ATG CCC AAC GCG 481 |
| Asp Ala I 65 | eu Thr Asn Ala Val Ala His Val Asp Asp Met Pro Asn Ala 70 75 |
| CTG TCC G | CC CTG AGC GAC CTG CAC GCG CAC AAG CTT CGG GTG GAC CCG 529 |
| Leu Ser A | la Leu Ser Asp Leu His Ala His Lys Leu Arg Val Asp Pro |

| SEQ.LS | ST ' | | Apri | 11 26 | 5, 19 | 993 | | | Page | e 1-1 | 12 | | | | | |
|---|--|---|---|---|----------------------------|---------------------------------|--|---|--------------------------------|-------------------------|---------------------------------------|---------------------------------------|----------------------------|--------------------------------|------|----|
| 80 | | | | 85 | | | | | 90 | , . | | | | 95 | | |
| GTC AA | AC TTC | AAG | CTC | CTA | AGC | CAC | TGC | CTG | CTG | GTG | ACC | CTG | GCC | GCC | 57 | 77 |
| Val As | n Phe | Lys | Leu 100 | Leu | Ser | His | Cys | Leu 105 | Leu | Val | Thr | Leu | Ala 110 | Ala | | |
| CAC CI | c ccc | GCC | GAG | TTC | ACC | CCT | GCG | GTG | CAC | GCC | TCC | CTG | GAC | AAG | 62 | 25 |
| His Le | eu Pro | Ala 115 | Glu | Phe | Thr | Pro | Ala 120 | Val | His | Ala | Ser | Leu 125 | Asp | Lys | | |
| TTC CT | G GCT | TCT | GTG | AGC | ACC | GTG | CTG | ACC | TCC | AAA | TAC | CGT | TAAC | GCTGGA | G 67 | 77 |
| Phe Le | eu Ala 130 | | Val | Ser | Thr | Val 135 | Leu | Thr | Ser | Lys | Tyr 140 | Arg | | | | |
| CCTCGG | STAGC | CGTT | CCTC | CT GO | CCCG | GTCG <i>I</i> | A CC | 70 |)9 | | | | | | | |
| (2) | NFORM | OITA | V FOR | R SEÇ | Q ID | NO:8 | 3: | | | | | | | | | |
| (i) | (. | B) : | LENGT | | 141 nino | ami acio | ino a 1 | cids | 3 | | | | | | | |
| | | | | | | | | | | | | | | | | |
| (ii) | М | OLEC | JLE 7 | TYPE: | : pro | oteir | 1 | | | | | | | | | |
| | M S | | | | _ | | | ID N | 10 : 8 : | ı | | | | | | |
| | s | EQUEI | NCE I Ala | DESCF | - Lys | ION: Thr | SEQ Asn | Val | Lys | Ala | | - | _ | _ | | |
| (ix) | S eu Ser | EQUEI Pro | NCE I Ala 5 | DESCI Asp | RIPT: | ION: Thr | SEQ Asn | Val 10 | Lys | Ala | | - | 15 | - | | |
| (ix) Val Le | S eu Ser .y Ala | EQUEN Pro His 20 Phe | NCE I Ala 5 Ala | DESCE Asp Gly | Lys Glu | ION: Thr Tyr | SEQ Asn Gly 25 | Val 10 Ala | Lys Glu | Ala Ala | Leu | Glu 30 | 15 Arg | Met | | |
| (ix) Val Le 1 Val Gl Phe Le | Seu Ser y Ala eu Ser 35 | Pro His 20 Phe | NCE I Ala 5 Ala Pro | Asp Gly | Lys Glu Thr | ION: Thr Tyr Lys 40 | SEQ Asn Gly 25 Thr | Val 10 Ala Tyr | Lys Glu Phe | Ala Ala Pro | Leu His 45 | Glu 30 Phe | 15 Arg Asp | Met Leu | | |
| (ix) Val Le 1 Val Gl Phe Le | Seu Ser Ly Ala eu Ser 35 | EQUENT Pro His 20 Phe | NCE I Ala 5 Ala Pro Ala | Asp Gly Thr | Lys Glu Thr Val | Thr Tyr Lys 40 Lys | SEQ Asn Gly 25 Thr | Val 10 Ala Tyr | Lys Glu Phe Gly | Ala Pro Lys 60 | Leu His 45 Lys | Glu 30 Phe Val | 15 Arg Asp Ala | Met Leu Asp | | |
| (ix) Val Le 1 Val Gl Phe Le Ser Hi | Seu Ser y Ala eu Ser 35 ds Gly eu Thr | Pro His 20 Phe Ser | NCE I Ala 5 Ala Pro Ala Ala | Asp Gly Thr Gln Val | Lys Glu Thr Val 55 | Thr Tyr Lys 40 Lys | SEQ Asn Gly 25 Thr Gly Val | Val 10 Ala Tyr His | Lys Glu Phe Gly Asp 75 | Ala Pro Lys 60 Met | Leu His 45 Lys Pro | Glu 30 Phe Val | Arg Asp Ala Ala | Met Leu Asp Leu 80 | | |
| (ix) Val Le 1 Val Gl Phe Le Ser Hi 5 Ala Le 65 | Seu Ser Y Ala Eu Ser 35 Ls Gly Eu Thr La Leu | EQUENT Pro His 20 Phe Ser Asn | NCE I Ala 5 Ala Pro Ala Ala Asp 85 | Gly Thr Gln Val 70 Leu | Lys Glu Thr Val 55 Ala | Thr Tyr Lys 40 Lys His | SEQ Asn Gly 25 Thr Gly Val | Val 10 Ala Tyr His Asp | Lys Glu Phe Gly Asp 75 Leu | Ala Pro Lys 60 Met Arg | Leu His 45 Lys Pro Val | Glu 30 Phe Val Asn | Asp Ala Ala Pro | Met Leu Asp Leu 80 Val | | |
| (ix) Val Le 1 Val Gl Phe Le Ser Hi 5 Ala Le 65 Ser Al | Seu Ser Y Ala Eu Ser 35 Ls Gly Eu Thr La Leu ne Lys | EQUENT Pro His 20 Phe Ser Asn Ser Leu 100 Glu | NCE I Ala 5 Ala Pro Ala Ala Asp 85 Leu | OESCR Asp Gly Thr Gln Val 70 Leu | Lys Glu Thr Val 55 Ala His | Thr Tyr Lys 40 Lys His Ala Cys | SEQ Asn Gly 25 Thr Gly Val His | Val 10 Ala Tyr His Asp Lys 90 Leu | Lys Glu Phe Gly Asp 75 Leu Val | Ala Pro Lys 60 Met Arg | Leu His 45 Lys Pro Val Leu | Glu 30 Phe Val Asn Asp | Arg Asp Ala Ala Pro 95 Ala | Met Leu Asp Leu 80 Val | | |

Leu Ala Ser Val Ser Thr Val Leu Thr Ser Lys Tyr Arg

SEQ.LST April 26, 1993 Page 1-13

135

(2) INFORMATION FOR SEQ ID NO:9:

130

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 743 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (vii) IMMEDIATE SOURCE:

35

- (B) CLONE: beta-hemoglobin
- (ix) FEATURE:
 - (A) NAME/KEY: transit_peptide (B) LOCATION: 26. .241

140

- (B) LOCATION: 26..241
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 245..685
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

| CTCGAGGGGA TCTGATCTTT CAAGAATGGC ACAAATTAAC AACATGGCAC AAGGGATACA | 60 |
|--|-----|
| AACCCTTAAT CCCAATTCCA ATTTCCATAA ACCCCAAGTT CCTAAATCTT CAAGTTTTCT | 120 |
| TGTTTTTGGA TCTAAAAAAC TGAAAAATTC AGCAAATTCT ATGTTGGTTT TGAAAAAAGA | 180 |
| TTCAATTTTT ATGCAAAAGT TTTGTTCCTT TAGGATTTCA GCAGGTGGTA GAGTTTCTTG | 240 |
| GATG GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG | 289 |
| Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp 1 5 10 15 | |
| GGC AAG GTG AAC GTG GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG | 337 |
| | |
| Gly Lys Val Asn Val Asp Glu Val Gly Glu Ala Leu Gly Arg Leu 20 25 30 | |

Leu Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp

CTG TCC ACT CCT GAT GCT GTT ATG GGC AAC CCT AAG GTG AAG GCT CAT

40

45

| SEQ | LST | , | | Apri | 11 26 | 5, 19 | 93 | | | Page | e 1-1 | L4 | | | | |
|-------------------|------------------------|-----------------------------------|--|--|--|--|---|-----------------------------------|-------------------|------------|------------|------------|------------|------------|---------------------|-------|
| Leu (| Ser | Thr 50 | Pro | Asp | Ala | Val | Met 55 | Gly | Asn | Pro | Ĺys | Val 60 | Lys | Ala | His | |
| GĢC | AAG | AAA | GTG | CTG | GGT | GCC | TTT | AGT | GAT | GGC | CTG | GCT | CAC | CTG | GAC | 481 |
| Gly | Lys 65 | Lys | Val | Leu | Gly | Ala 70 | Phe | Ser | Asp | Gly | Leu 75 | Ala | His | Leu | Asp | |
| AAC | CTC | AAG | GGC | ACC | TTT | GCC | ACC | CTG | AGT | GAG | CTG | CAC | TGT | GAC | AAG | 529 |
| Asn 80 | Leu | Lys | Gly | Thr | Phe 85 | Ala | Thr | Leu | Ser | Glu 90 | Leu | His | Суз | Asp | L ys . 95 | |
| CTG | CAC | GTG | GAT | CCT | GAG | AGC | TTC | AGG | CTC | CTA | GGC | AAC | GTG | CTG | GTC | 577 |
| Leu | His | Val | Asp | Pro 100 | Glu | Ser | Phe | Arg | Leu 105 | Leu | Gly | Asn | Val | Leu 110 | Val | |
| TGT | GTG | CTG | GCG | CAT | CAC | TTT | GGC | AAA | GAA | TTC | ACC | CCA | CCA | GTG | CAG | 625 |
| Cys | Val | Leu | Ala 115 | His | His | Phe | Gly | Lys 120 | Glu | Phe | Thr | Pro | Pro 125 | Val | Gln | |
| GCT | GCC | TAT | CAG | AAA | GTG | GTG | GCT | GGT | GTG | GCT | AAT | GCC | CTG | GCC | CAC | 673 |
| Ala | Ala | Tyr 130 | Gln | Lys | Val | Val | Ala 135 | Gly | Val | Ala | Asn | Ala 140 | Leu | Ala | His | |
| AAG | TAT | CAC | TAAC | CTCC | CT T | TCTT | GCT | T C | CAATT | TCT | A TTA | AAAGO | TTC | | | 722 |
| Lys | Tyr 145 | His | | | | | | | | | | | | | | |
| CTT | rgtgo | GG 7 | CGAC | GTC | A C | | | | | | | | | | | 743 |
| | | | | | | | | | | | | | | | • | , 15 |
| | | | | | | | | | | | | | | | | 7 3 3 |
| (2) | INI | FORM | OITA | 1 FOE | R SEÇ |) ID | NO: | 10: | | | | | | | | 743 |
| (5 | L) | SI (1 (I | EQUEN A) LI B) T | NCE (ENGTH FYPE: FOPOI | CHARA H: 14 : ami | ACTER 6 an ino a 1 lir | RIST: mino acid near | ICS: acio | ls | | | | | | | 743 |
| (i: | L) | SI (1 (I MO | EQUEN A) LI B) T D) T | NCE (ENGTH TYPE: TOPOI JLE T | CHARA H: 14 : ami LOGY: | ACTER 6 an 1no a 1 lir | RISTI mino acid near oteir | ICS: acio | | √o∙ 1 | 10. | | | | | 713 |
| (ii (xi | i) i) i) | SI (I (I MC | EQUEI A) LI B) T OLECU | NCE (ENGTHE STOPOLICE TOPOLICE | CHARA H: 14 : ami LOGY: CYPE: | ACTER 16 and 1no a 1 lir 2 pro | RIST: mino acid near oteir | ICS: acid | ID 1 | | | Δla | I.eu | Trn | Glv | , 13 |
| (ii (xi | i) i) i) | SI (I (I MC | EQUEI A) LI B) T OLECU | NCE (ENGTHE STOPOLICE TOPOLICE | CHARA H: 14 : ami LOGY: CYPE: | ACTER 16 and 1no a 1 lir 2 pro | RIST: mino acid near oteir | ICS: acio | ID 1 | | | Ala | Leu | Trp 15 | Gly | , 13 |
| (i: (x: Val | i) i) i) His | SI (I (I MC SI Leu | EQUEN 3) TO DLECT EQUEN | NCE (ENGTH TYPE: TOPOI JLE T NCE I Pro 5 | CHARA H: 14 : ami LOGY: TYPE: DESCR | ACTER 16 an 1no a 1 lir 2 pro RIPTI | RISTI mino acid lear bteir ION: | ICS: acid | ID N Ala 10 | Val | Thr | | | 15 | _ | |
| (iii (xi Val 1 | i) i) His Val | SI (I (I MC SI Leu | EQUENCY OF THE TAIL OF THE TAI | NCE (ENGTH TYPE: TOPOI JLE T NCE I Pro 5 | CHARA H: 14 : ami LOGY: TYPE: DESCR | ACTER 6 and | RISTI nino acid lear otein ION: Lys | ICS: acid seq Seq Gly | ID NAla 10 | Val Ala | Thr Leu | Gly | Arg 30 | 15 Leu | Leu | |

Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His Gly 50 55 60

Lys Lys Val Leu Gly Ala Phe Ser Asp Gly Leu Ala His Leu Asp Asn

Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys Leu 90

His Val Asp Pro Glu Ser Phe Arg Leu Leu Gly Asn Val Leu Val Cys

Val Leu Ala His His Phe Gly Lys Glu Phe Thr Pro Pro Val Gln Ala 115

Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His Lys 130 135

Tyr His 145

SEQ.LST

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - LENGTH: 17 amino acids (A)
 - TYPE: amino acid (B)
 - TOPOLOGY: linear (D)
- (ii) MOLECULE TYPE: peptide
- (V) FRAGMENT TYPE: N-terminal
- (vi) ORIGINAL SOURCE:
 - ORGANISM: alkalophilic Bacillus sp.
 - (B) STRAIN: 38-2
- IMMEDIATE SOURCE: (vii)
 - CLONE: beta-cyclodextrin
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Ala Pro Asp Thr Ser Val Ser Asn Lys Gln Asn Phe Ser Thr Asp Val 15

Ile